

TECNOLOGIA EM BANCO DE DADOS - 2025

Análise de Dados sobre Mortes em Excesso Associadas à COVID-19: Estudo Aplicado de Ciência de Dados

DISCIPLINA: PROJETO APLICADO I - BANCO DE DADOS - 2025.2

GRUPO 11:

DANIEL SILVA SOUZA FILHO

VITOR CLEMES NOVAKOSKI

CAIO YAMAOKA BARBOSA

NAIARA CRISTINA GUEDES INOCENCIO

<https://github.com/VitorNovakoski/Grupo11-WWO.git>

SUMÁRIO

Glossário.....	3
Objetivo do Estudo.....	3
Apresentação da Empresa.....	4
Problema do Estudo.....	5
Metadados.....	6
Análise Exploratória de Dados.....	7
Referências:.....	9

Análise de Dados sobre Mortes em Excesso Associadas à COVID-19: Estudo Aplicado de Ciência de Dados

DANIEL SILVA SOUZA FILHO

VITOR CLEMES NOVAKOSKI

CAIO YAMAOKA BARBOSA

NAIARA CRISTINA GUEDES INOCENCIO

Glossário

OMS (Organização Mundial da Saúde) – Agência das Nações Unidas especializada em saúde pública internacional, responsável pela coleta e divulgação dos dados utilizados neste estudo.

Excesso de mortes – Diferença entre o número de mortes observadas e o número esperado em condições normais (sem pandemia).

Dataset – Conjunto estruturado de dados disponibilizado para análise.

Data Science (Ciência de Dados) – Campo interdisciplinar que combina estatística, computação e conhecimento de domínio para extrair informações relevantes de dados.

EDA (Exploratory Data Analysis) – Análise Exploratória de Dados; processo inicial de investigação de um conjunto de dados para descobrir padrões e anomalias.

Objetivo do Estudo

O objetivo deste estudo é analisar os dados de mortalidade em excesso associados à COVID-19, disponibilizados pela Organização Mundial da Saúde (OMS), com o intuito de identificar padrões, tendências e anomalias regionais e temporais.

A análise visa compreender o impacto real da pandemia em diferentes países e propor soluções analíticas que permitam monitorar, comparar e comunicar os resultados de forma clara e acessível, tanto a públicos técnicos quanto não técnicos.

Apresentação da Empresa

- **Nome da organização:**

- Organização Mundial da Saúde (World Health Organization – WHO / OMS)

- **Missão, Visão e Valores**

- **Missão:** Promover a saúde, manter o mundo seguro e servir aos vulneráveis.
- **Visão:** Atingir o mais alto nível possível de saúde para todas as pessoas.
- **Valores:** Comprometimento com a ética, equidade, colaboração científica e transparência nas ações globais de saúde.

- **Segmento de atuação**

A OMS atua no segmento de saúde pública internacional, coordenando esforços de vigilância epidemiológica, campanhas de vacinação, prevenção de doenças e desenvolvimento de políticas sanitárias.

O foco deste estudo recai sobre o segmento de dados epidemiológicos e vigilância global de mortalidade, parte integrante das ações de monitoramento da COVID-19.

- **Market Share e posicionamento**

Como agência líder em saúde global, a OMS é a principal referência mundial na coleta e padronização de dados sanitários, servindo de base para políticas públicas, pesquisas acadêmicas e decisões governamentais.

- **Número de colaboradores**

A organização conta com aproximadamente 8.000 colaboradores distribuídos em mais de 150 países e escritórios regionais.

- **Iniciativas em Data Science**

A OMS possui iniciativas robustas em Ciência de Dados e Big Data, utilizando plataformas como o WHO Data Hub e sistemas de Open Data para disponibilizar conjuntos de dados sobre saúde, mortalidade, vacinação, doenças transmissíveis e fatores ambientais.

- **Trabalhos em destaque**

- Sistema Global Health Observatory (GHO).
- Projeto COVID-19 Dashboard, que monitora casos e óbitos em tempo real.
- Estudo Global Excess Deaths Associated with COVID-19, objeto desta pesquisa.

Problema do Estudo

- **Contexto e motivação**

Durante a pandemia da COVID-19, muitos países apresentaram discrepâncias entre o número de mortes relatadas e o número real estimado.

Esse fenômeno gerou a necessidade de estimar mortes em excesso, abrangendo tanto os óbitos diretos pela doença quanto os indiretos, decorrentes da sobrecarga dos sistemas de saúde e outros fatores socioeconômicos.

- **Perguntas orientadoras**

- O que explica a variação de mortes em excesso entre países e regiões?
- Existem padrões sazonais ou geográficos evidentes nos dados?
- Quais países apresentaram subnotificação mais expressiva?
- O impacto da pandemia foi uniforme ao longo do tempo ou apresentou picos específicos?
- É possível estimar tendências futuras a partir do comportamento observado?

- **Abordagem de pensamento computacional**

- **Decomposição:** O problema é dividido em partes menores — análise temporal, análise geográfica e análise de qualidade dos dados.
- **Reconhecimento de padrões:** Identificação de comportamentos similares entre países e períodos.
- **Abstração:** Foco nas variáveis mais relevantes (excesso, observadas, esperadas), ignorando detalhes que não influenciam as tendências gerais.
- **Algoritmização:** Implementação de scripts em Python para automatizar o processamento, cálculo e visualização dos resultados.

Metadados

- **Descrição geral**

O dataset utilizado foi obtido no portal oficial da OMS: <https://www.who.int/data/sets/global-excess-deaths-associated-with-covid-19-modelled-estimates>

Item	Descrição
Tipo de arquivo	CSV e XLSX
Origem dos dados	Dados abertos da Organização Mundial da Saúde (OMS)
Sensibilidade	Dados não sensíveis, agregados por país/região
Validade	Coleta e estimativas referentes a 2020–2021, com comparativo pré-pandemia
Proprietário do dado	Organização Mundial da Saúde (World Health Organization – WHO)
Restrições de uso	Dados de uso público, sujeitos à citação obrigatória da fonte. Sem dados pessoais, em conformidade com a LGPD
Informações adicionais	Estimativas obtidas por modelagem estatística e registros civis nacionais
Principais atributos	País/Região, Ano/Mês, Mortes Observadas, Mortes Esperadas, Excesso de Mortes, Intervalo de Confiança
Tipo de dado	Quantitativo contínuo e categórico
Formato	Estruturado, tabular, delimitado por vírgula (CSV)

Análise Exploratória de Dados

Os scripts desenvolvidos em Python realizam o tratamento e exploração inicial dos dados, produzindo gráficos, tabelas e rankings que apoiam a compreensão das tendências globais.

- Principais resultados gerados:
 - **Histograma** de distribuição das mortes em excesso.
 - **Série temporal global** (soma mensal).

- **Ranking Top 15 e Bottom 15 países** por excesso acumulado.
- **Cobertura de dados por país** (número de registros, período, faltantes).
- **Perfis individuais de países** (Brasil, Índia, EUA, entre outros).

Essas análises iniciais servirão de base para as próximas etapas do projeto, especialmente a Etapa 3 – Storytelling, onde os resultados serão comunicados por meio de narrativas visuais e interpretativas.

Referências:

WORLD HEALTH ORGANIZATION. Global excess deaths associated with COVID-19 (modelled estimates). WHO Data. Disponível em: <https://www.who.int/data/sets/global-excess-deaths-associated-with-covid-19-modelled-estimates>. Acesso em: [15/out/2025].

GOLDSCHMIDT, Ronaldo; BEZERRA, Eduardo. Data mining: conceitos, técnicas, algoritmos, orientações e aplicações. 2. ed. Rio de Janeiro: Elsevier, 2015.

CAPPRA, Ricardo. Rastreável: redes, vírus, dados e tecnologias para proteger e vigiar a sociedade. São Paulo: Actual, 2021.